



Hyperparameter Optimisation service at ATLAS

Alkaid Cheng, Wen Guan, Fa-Hui Lin, Fernando Barreiro Megino, Tadashi Maeno, Paul Nilsson, Torre Wenaus, <u>Rui Zhang</u>

University of Wisconsin-Madison

IML Machine Learning

Mar 16, 2021

Contents

- Introduction to the topic
- intelligent Data Delivery Service (iDDS)
- Hyperparameter optimisation using iDDS on the ATLAS Grid
 - The normal HPO workflow
 - Ingredients of the workflow
 - The segmented HPO workflow
- Distributed learning on heterogeneous resources
 - HPC/Summit and its challenges
 - Commercial Cloud
- Visualisation support

Introduction: HPO service at ATLAS

- The goal of the project is to provide an HPO service to ATLAS users for ML
 - Minimal user code adaption
 - Support for advanced search algorithms in addition to the traditional grid or random search algorithms
 - Reuse ATLAS production and distributed system (PanDA) no reinventing the wheel
 - Visualisation of results
- Single-function-call pattern for HPO
 - Computing resources are managed behind the scene
 - Not suitable since ATLAS has its own resource management
- Ask-and-tell pattern for HPO
 - Decoupled optimisation+sampling from training in space-time
 - Purely point searching, no resource management
 - We go in this way

"The ask-and-tell pattern"
while ~ opt.stop
 x = ask(opt)

```
y = f(x)
opt = tell(opt, x, y)
end
```

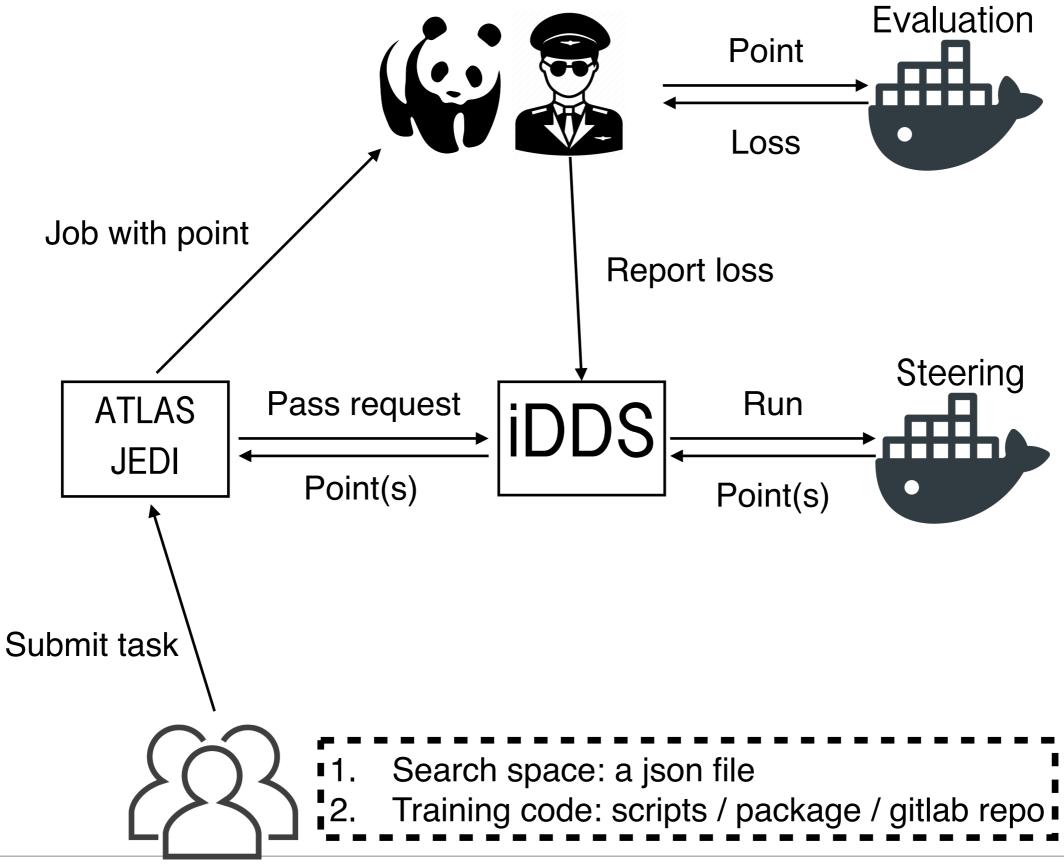
The intelligent Data Delivery Service (iDDS)

- <u>iDDS</u> (joint ATLAS/IRIS-HEP project) is designed to intelligently deliver needed data/jobs to workflows in a fine-grained way
 - Data Carousel: jobs start when its own input is ready, no wait for the full dataset to be transferred
- **iDDS** Request • DAG (directed acyclic Requester Head graph) based workflow: successive jobs start off Conducto Carrier Clerk Transformer ÷ once all dependent jobs are done DDMS Scheduler backend for subsequent HPO is a series of tasks with Requests Workflow work step Transformation Backend Schedule' Work decision-making in between Transform Original Work collection Deliver Consumer Work —- another paradigm New Storage collection Storage

В

А

Overview of the HPO workflow



Ingredients of the workflow

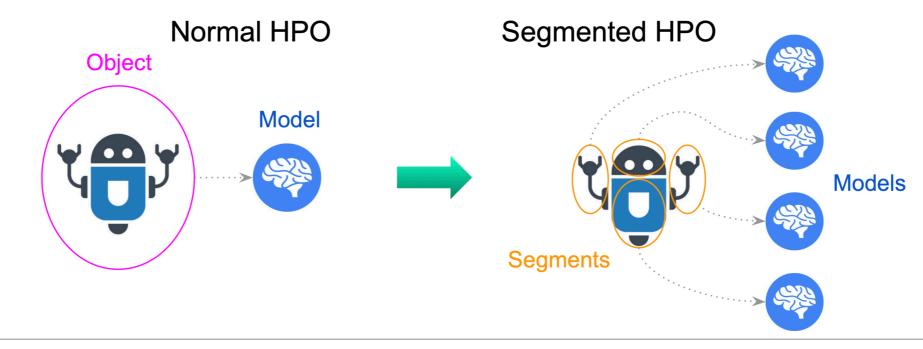
- Two containers to fulfil the loop:
 - SteeringContainer optimisation at iDDS servers
 - Generate next HP points with customised method
 - A wide range of HPO methods are supported
 - EvaluationContainer training at Grid (GPU) sites
 - Submodule payload contains model definition, training scripts (user specific)
- Checkpointing:
 - Periodically upload checkpoints to Grid
 - Download the checkpoint when the same job is retrying
 - Resume training if checkpoint is found

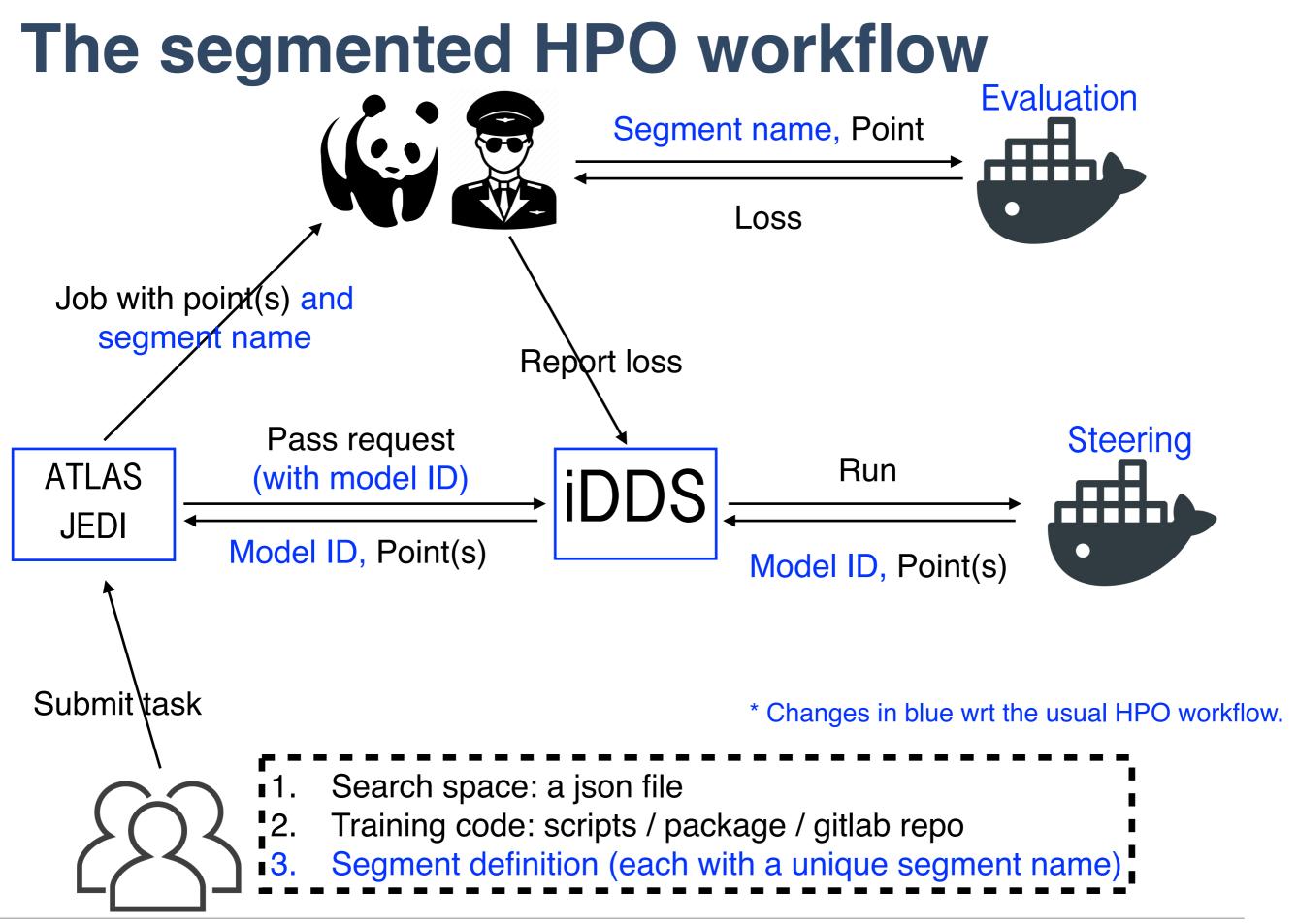




Segmented HPO - Why do we need it

- Some machine learning payloads have similar architecture of models targeting different physics regions/objects
 - Essentially multiple models formed with different training datasets
 - Once the amount is large, bookkeeping is challenging
- A real ATLAS example: FastCaloGAN, a calorimeter image generation model
 - 300 GANs = 3 PIDs x 100 η slices
 - 300 individual tasks in the usual workflow
- ✤ Now this can be done with Segmented HPO



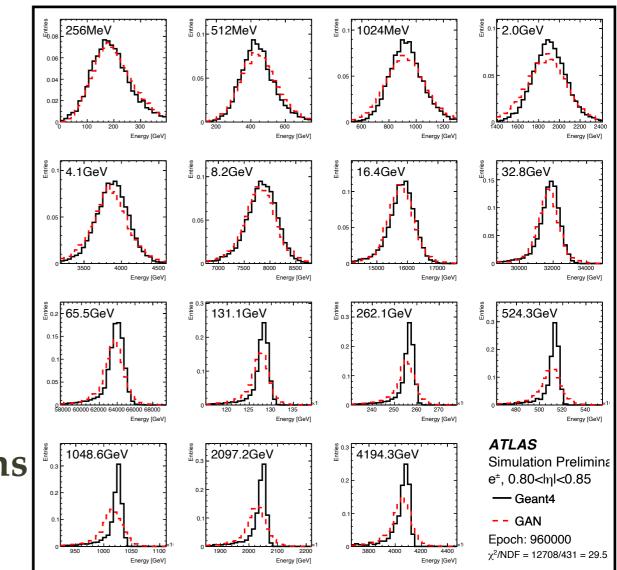


Features and test results

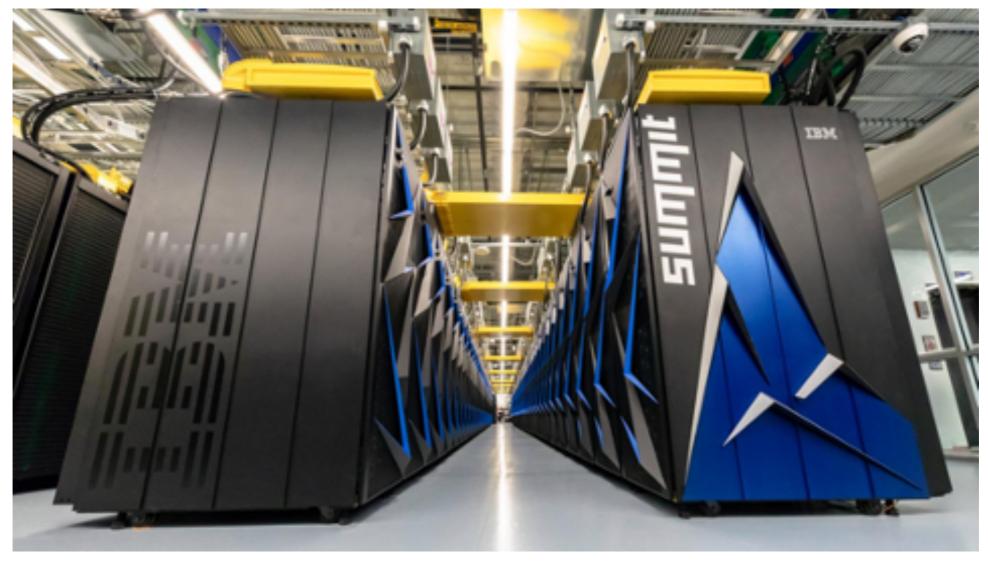
- Features:
 - Support both normal and segmented HPO tasks
 - An argument to fill when submitting
 - Support separating input for each segment
 - To reduce the load of the sites
- Tested with 15 GANs
 - 3 particle types $\times 5 \eta$ slices
- Plot on the right is from a 10K epochs
 job from the BNL GPU site

Input dataset (PID 22, 0 < $|\eta|$ < 0.05): pid22_eta_0_5.v02.tar

pid22_eta_0_5.v02.tar 🍙 🙋 🔗	input	ready
binning.xml 🚡 🖄 🔗	input	ready
user.zhangr.seg_gpuBNL.bd6dda1e-8db4-483b-b4f9-3d04dee8b340.log.24393491.000010.log.tgz 🍸 🧔	log	ready
photons_0_5.24393491.metrics.000010.tgz_1 👔 🚳	output	ready
photons_0_5.24393491.metrics.000010.tgz_0 👔 🔅	output	ready
pseudo_lfn	pseudo_input	unknown



HPCs as GPU resources

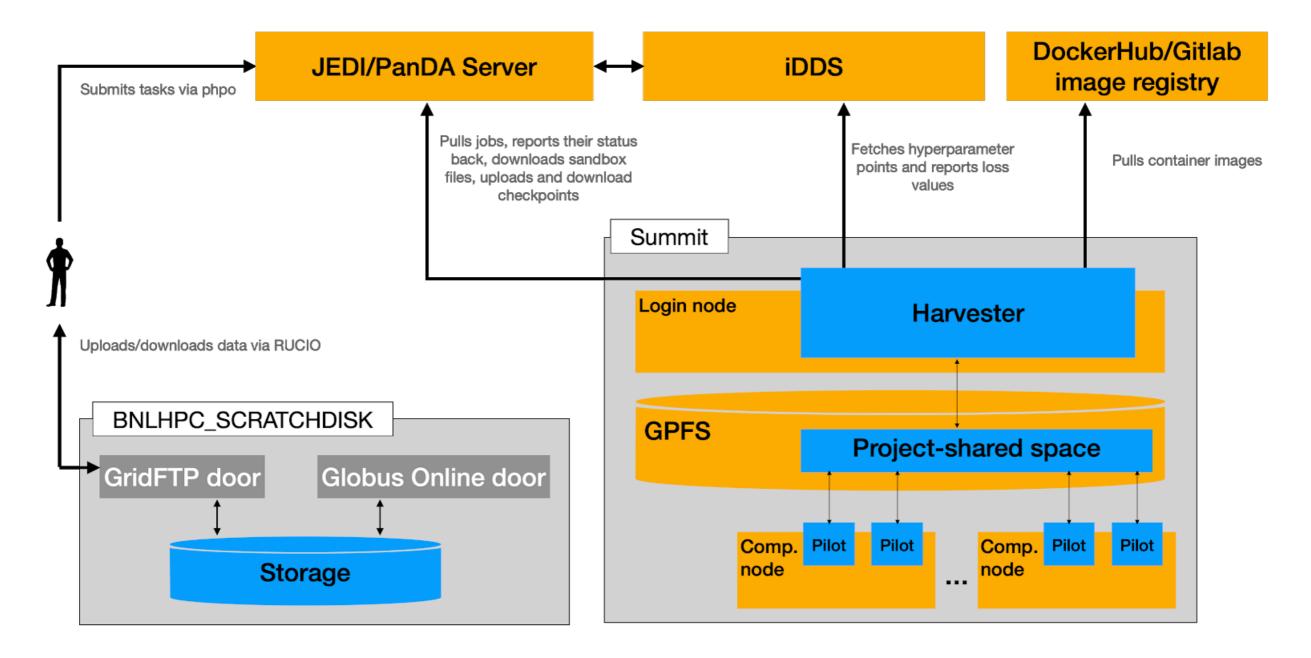


- Summit as an example
 - 4608 computer nodes
 - 2 Processors x 22 cores / node
 - 6 V100 GPUs / node
 - Wonderful workstation for ML/HPO

- Challenges
 - Short wall time
 - Standard Grid services and workflows unavailable or suboptimal

Integration map on Summit

- Harvester is a key component in the HPC environment
 - Need to connect it with JEDI, iDDS and image repository



Challenges on Summit / HPC

- Various issues need to be addressed for HPCs, which are different among HPCs. For Summit, there are following striking factors:
 - 1) PowerPC architecture a non-x86 architecture
 - Encounter different compilers for PowerPC 😐
 - Docker image is architecture-dependent; not straightforward to provide for a production service 😕
 - 2) Short walltime 2h for each job*
 - Checkpointing is almost always needed 😐
 - 3) No network inbound / outbound connectivity of worker nodes
 - Requires all demands to be downloaded in advance
 - This hurts a lot for many payloads that are designed without this limitation 🙁

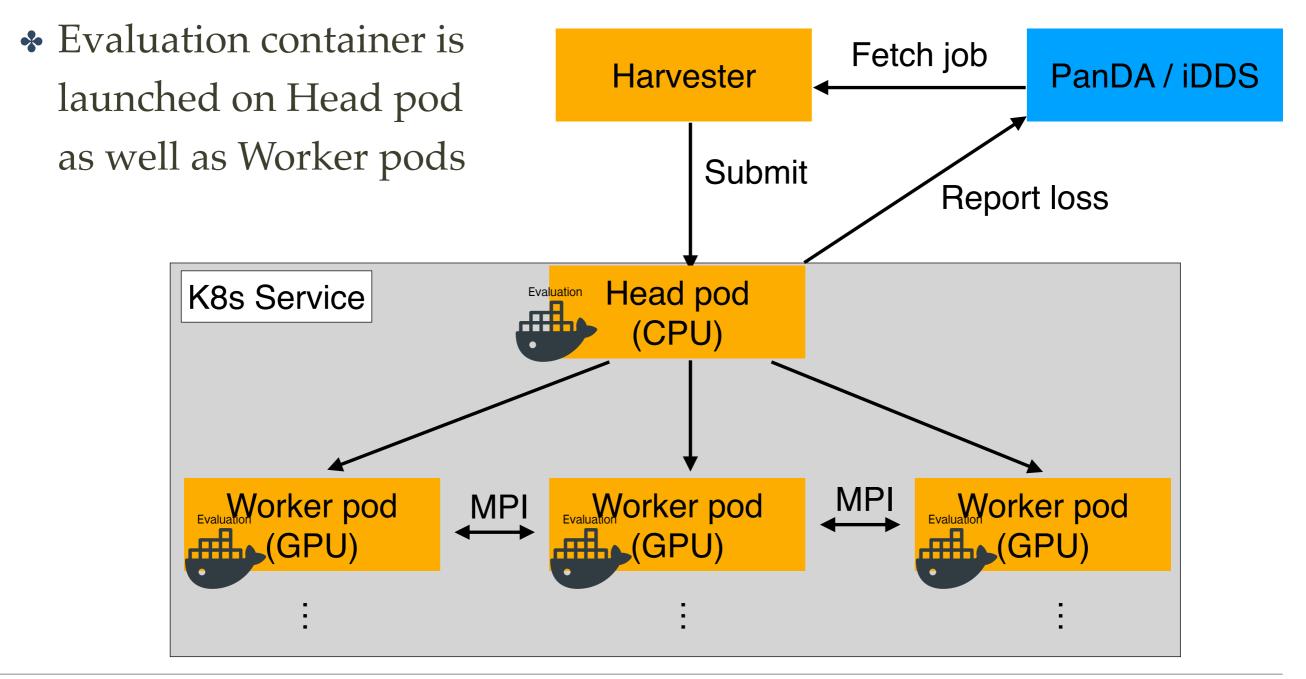
*1-45 nodes: 2h, 46-91 nodes: 6h, 92-921 nodes: 12h, 922-4608 nodes: 24h

Distributed training on commercial cloud

- Commercial cloud is one of the best places for distributed training
 - GPUs on the grid are mostly for single-GPU training
 - HPCs have many development and operational challenges
- So far in the R&D phase
 - Big investment for pledged GPU would be a bit risky since there are not many real use cases now
 - Should be prepared since distributed training is quite popular outside of HEP
- Horovod is currently being tested
 - A useable Evaluation container is created, to be tested with multi-GPU resources.
 - Open to support other distributed solutions, e.g. <u>DASK</u>, <u>Ray</u>

(Simplified) Integration map on Amazon K8s

- Horovodrun (MPI launcher) runs on the Head pod
- Number of Worker pods is scalable by K8s



Visualisation support

- A visualisation tool MLflow is turned on in EvaluationContainer
 - Useful for offline visualisation as it is part of the output
- An α -version of the tool also integrated into PanDA Monitoring system
 - Fetch output from Evaluation container (training job) and spin-up an MLFlow container to display results
 - Extendable to other nice visualisation tools (Neptune, WandB, Tensorboard, etc.) if outputs match the visualisation backend or if additional conversion step is implemented

■ <	> 0	🔒 big	panda.c	ern.ch		Ç				ŵ +		• • •		>	O			🔒 bigp	anda.cern.	.ch		Ś		(⊕	+
Experiments	Models									GitHub Docs		validation_spl	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	my-experiment											validation_step	s 12	195	97	48	6	97	24	24	6	97	195	48	195	19
+ <												workers	1	1	1	1	1	1	1	1	1	1	1	1	1	1
eriments	Experiment ID: 1	Artifact L	cation: file	:///srv/workDir/mlrun	is/1	. 11						. /	(120	201	17/									
∠ 🗓				n	TTDS	5.77	010	Dal	nda	a.ce	rn.cł	17mi	τιον	NIZ	426	129.	137					Metrics				
ent 🖉 🔟	✓ Notes					•//						accuracy let	0.812	0.812	0.777	0.81	0.80	0.821	0.815	0.812	0.651	0.82	0.818	0.508	0.812	C
	None											loss 🗠	0.4	0.405	0.452	0.403	0.415	0.387	0.399	0.408	0.609	0.389	0.391	0.693	0.404	0.
												stopped_epoch	· 0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Search Q metrics.rmse < 1 and para Runs:						0) \Xi Filter	Search	h Clear				0.000	0.017	-	0.005	0.000	0.005	-	0.770	0.007	0.007	0.555	0.000	-
												val_accuracy l	0.826	0.823	0.817	0.83	0.825	0.832	0.825	0.826	0.773	0.827	0.827	0.555	0.823	C
	Showing 20 matching runs Compare	Delete Download CSV						=		Columns		val_loss 🗠	0.381	0.386	0.399	0.371	0.381	0.37	0.383	0.381	0.483	0.384	0.379	0.691	0.385	C
					Parame	ters >		Metrics	>																	
	Start Time Run Na	ame User Source	Version	Models	baselin		siz class w	veil accuracy		stopped_e		A	0													
	○ 2021-02-20 10:00 -		88afbd	s keras	None	None	None	0.813	0.4	o o		Scatter Plot	Contour Plot	t Paralle	I Coordinates P	lot										
	□		88afbd	a keras	None	None	None	0.812	0.405	0		Parameters:		ort me	nen ultioro net siamsorada	monitor shufflet learn)	no materietateris per enm	wittion service	le weightropping workers	hatch statisfiation stant	decalmitial entroph quer	uerbiss weightig deltar	nt ensiloent hetava%datii	estfrenhest we hriddik e	ation batchealzioss	
		atlprd56 🗆 TestRu								0		Parameters:		opt_tas	njmultipro apt sjøgisgradi		90 0.990001718.0000D	01.00000 1.00000	le_weightpochs workers	195.000001.0	0000 1.00000 11.000	000 1.00000	0.00000 1.09890 1.10		ation_batch yalzdoss 0.69116	
	○ 2021-02-20 09:28 - ○ 2021-02-20 09:02 - ○ 2021-02-20 09:02 - ○ 2021-02-20 07:11 -	atlprd56	88afbd	keras	None	None	None	0.812	0.405	0 0 0			sing×	opt_base	nymultipro apt;jeg isgradi		90 0.990001718.00000	00 1.00000 1.00000	2.20000 1.10000	195.000001.0			0.00000 1.09890 1.10		0.69116	-
	○ 2021-02-20 09:29 - ○ 2021-02-20 09:02 - ○ 2021-02-20 07:11 - ○ 2021-02-20 06:45 -	atlprd56	88afbd 88afbd 88afbd 88afbd	⊵ keras ⊵ keras	None None None None	None None	None None	0.812 0.777 0.81 0.802	0.405 0.452 0.403 0.415	0 0 0 0		opt_name× use_multiproces opt_amsgrad×	monitor×	opt_max	ngmultipro api gjøgisgradt	0.02500	90 0.990001718.00000	00 1.00000 1.00000	2.20000 1.10000	195.000001.0	0000 1.00000 11.000	000 1.00000	0.00000 1.09890 1.10			0.6
	○ 2021-02-20 09:25 - ○ 2021-02-20 09:02 - ○ 2021-02-20 07:11 - ○ 2021-02-20 06:45 - ○ 2021-02-20 06:45 -	atlprd56	88afbd 88afbd 88afbd 88afbd 88afbd	 keras keras keras keras keras keras keras 	None None None None None	None None None None	None None None None	0.812 0.777 0.81 0.802 0.821	0.405 0.452 0.403 0.415 0.387	0 0 0 0 0	L	opt_name× use_multiproces opt_amsgrad× shuffle× opt_	monitor×	opt_max	njimultipro apt ijagnsgradn	0.02500	90 0.990001718.00000 98000 - 51.0000 96669900.00090 -	01.00000 1.00000	2.20000 1.10000	195.000001.0	1.0000 1.00000 11.000	000 1.00000 1.00000 -0.000	0.00000 1.09890 1.10		0.69116	0.6
		atlprd56 TestRu	88afbd 88afbd 88afbd 88afbd 88afbd 88afbd	 keras keras keras keras keras keras keras keras keras 	None None None None None None	None None None None None	None None None None None	0.812 0.777 0.81 0.802 0.821 0.815	0.405 0.452 0.403 0.415 0.387 0.399	0 0 0 0 0 0 0	L	opt_name× use_multiproces opt_amsgrad× shuffle× opt_ opt_beta_1×	monitor× earning_rate×		ngmultiproæpt _e jøgrsgradt	0.02500-0 0.02500-0 0.02000-0	90 0.990001718.00000 98000 - 11 0000 36669900.00090 - 10.5000 - 10.5000 9000.90000 -	00 -0.50000 -	2.20000 1.10000 2.20000 1.10000 2.15000 -	195.00000 1.0	1.0000 1.00000 11.000	000 1.00000 1.00000 -0.000	0.00000 1.09890 1.10		0.69116	0.6
	○ 2021-02-20 09:25 - ○ 2021-02-20 09:02 - ○ 2021-02-20 07:11 - ○ 2021-02-20 07:11 - ○ 2021-02-20 07:11 - ○ 2021-02-20 07:11 - ○ 2021-02-20 07:14 - ○ 2021-02-20 06:45 - ○ 2021-02-20 04:45 - ○ 2021-02-20 04:45 - ○ 2021-02-20 04:45 -	atlprd56	88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd	keras keras keras keras keras keras keras keras	None None None None None None	None None None None None None	None None None None None None	0.812 0.777 0.81 0.802 0.821 0.815 0.812	0.405 0.452 0.403 0.415 0.387 0.399 0.408	0 0 0 0 0 0 0 0 0	L	opt_name× use_multiproces opt_amsgrad× shuffle× opt_ opt_beta_1× steps_per_epoc	monitor× earning_rate×	K		0.02500 0.02500 0.02000 0.02000	90 0.990001718.00000 98000 - 11.0000 36000 - 10.5000 - 194000 - 10.5000 5000.00000 - 5000.00000 - 522000	DL.00000 1.00000 00 -1.00000 - 00 -0.50000 -	2.20000 1.10000 2.20000 1.10000 2.15000 - 2.15000 - 2.05000 - 2.05000 -	195.00000 1.0	1.0000 1.00000 11.000	000 1.00000 1.00000 -0.000	0.00000 1.09890 1.10		0.69116	0.6
		atlprd56 — TestRu	88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd	© keras © keras © keras © keras © keras © keras © keras © keras	None None None None None None None	None None None None None None None	None None None None None None None	0.812 0.777 0.81 0.802 0.821 0.815 0.812 0.651	0.405 0.452 0.403 0.415 0.387 0.399 0.408 0.609	0 0 0 0 0 0 0 0 0 0	L	opt_name× use_multiproces opt_amsgrad× shuffle× opt_ opt_beta_1×	monitor× earning_rate×	K	nymultiproæptgjøgnsgradt – False — False – val_log	0.02500 0.02500 0.02000 0.02000	90 0.990001718.00000 98000 - 11 0000 36669900.00090 - 10.5000 - 10.5000 9000.90000 -	DL.00000 1.00000 00 -1.00000 - 00 -0.50000 -	2.20000 1.10000 2.20000 1.10000 7 2.15000 - 2.10000 -1.05000 -	195.00000 1.0	1.0000 1.00000 11.000	000 1.00000 1.00000 -0.000	0.00000 1.09890 1.10		0.69116 0.65000 - 0.60000 -	0.6
		atlprd56 — TestRu atlprd56 — TestRu	88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd	© keras © keras © keras © keras © keras © keras © keras © keras © keras	None None None None None None None None	None None None None None None None None	None None None None None None None None	0.812 0.777 0.81 0.802 0.821 0.815 0.812 0.651 0.82	0.405 0.452 0.403 0.415 0.387 0.399 0.408 0.609 0.389	0 0 0 0 0 0 0 0 0 0 0 0	L	opt_name× use_multiproces opt_amsgrad× shuffle× opt_ opt_beta_1× steps_per_epoc validation_split>	monitor× earning_rate× n× patience> epochs×	K		0.02500 0.02500 0.02000 0.02000	90 0.990001718.00000 98000 - 11.0000 36000 - 10.5000 - 194000 - 10.5000 5000.00000 - 5000.00000 - 522000	BL.00000 1.00000 100 -1.00000 - 00 -0.50000 - 00 -0.50000 - None	2.20000 1.10000 2.20000 1.10000 2.15000 - 2.15000 - 2.05000 - 2.05000 -	195.00000 1.0	1.0000 1.00000 11.000	000 1.00000 1.00000 -0.000	0.00000 1.09890 1.10		0.69116 0.65000 - 0.60000 -	0.6
	○ 2021-02-20 09:25 - ○ 2021-02-20 09:02 - ○ 2021-02-20 07:11 - ○ 2021-02-20 07:11 - ○ 2021-02-20 07:11 - ○ 2021-02-20 07:11 - ○ 2021-02-20 07:11 - ○ 2021-02-20 07:14 - ○ 2021-02-20 04:25 - ○ 2021-02-20 04:25 - ○ 2021-02-20 01:26 - ○ 2021-02-20 01:56 - ○ 2021-02-20 01:57 -	atlprd56 _ TestRu atlprd56 _ TestRu	88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 4ae72d	B keras	None None None None None None None None	None None None None None None None None	None None None None None None None None	0.812 0.777 0.81 0.802 0.821 0.815 0.815 0.812 0.651 0.82 0.818	0.405 0.452 0.403 0.415 0.387 0.399 0.408 0.609 0.389 0.391	0 0 0 0 0 0 0 0 0 0 0 0 0 0	l	opt_name× use_multiproces opt_amsgrad× shuffle× opt_ opt_beta_1× steps_per_epoc validation_split× sample_weight>	monitor× earning_rate× n× patience> epochs× ch_size×	< Adam —		0.02500 0.02500 0.02000 0.02000	90 0.990081718.0000 98000 11.000 9666990.00090 9000.00000 9000.00000 9000.00000 90000000 900000 900000 900000 900000 900000 900000 9000000 9000000 900000000	BL.00001 1.00000 00 -1.00000 00 -0.50000 -	2.2000 1.10000 2.20000 -1.10000 2.15000 - 2.10000 -1.05000 - 2.05000 - 2.05000 - 1.95000 -	195.00001.0 1.0000 150.0000 0.5000	0000 1.0000 11.000 1.00000 1.00000 7 0.50000 10.50000 - 0.50000 10.50000 -	000 1.00000 1.00000 -0.000 0.50000 -0.000	0.00000 1.09690 1.10 00 -1.10000 - 00 -1.05000 -1.05000 - 00 -1.05000 -1.05000 -		0.69116 0.65000 - 0.60000 -	0.6
	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	atiprd56 - TestRu atiprd56 - TestRu	88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 4ae72d	B keras	None None None None None None None None	None None None None None None None None	None None None None None None None None	0.812 0.777 0.81 0.802 0.821 0.815 0.815 0.812 0.651 0.82 0.818 0.508	0.405 0.452 0.403 0.415 0.387 0.399 0.408 0.609 0.389 0.391 0.693	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	l	opt_name× use_multiproces opt_amsgrad× shuffle× opt_ opt_beta_1× steps_per_epoc validation_split> sample_weight) workers× bat validation_steps initial_epoch×	monitor× earning_rate× n× patience> epochs× ch_size× x opt_decay	< Adam —		0.02500 0.02500 0.02000 0.02000	90 0.990081718.0000 98000 11.000 9666990.00090 9000.00000 9000.00000 9000.00000 90000000 900000 900000 900000 900000 900000 900000 9000000 9000000 900000000	BL.00001 1.00000 00 -1.00000 00 -0.50000 -	2.2000 1.10000 2.20000 1.10000 2.15000 - 2.15000 - 2.05000 - 1.95000 - 1.95000 - 1.95000 -	195.00000 1.0 1.00000	0000 1.0000 11.000 1.00000 1.00000 7 0.50000 10.50000 - 0.50000 10.50000 -	000 1.00000 1.00000 -0.000 0.50000 -0.000	0.00000 1.09890 1.10		0.69116 0.65000 - 0.60000 -	0.6
	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	atlprd56 - TestRu atlprd56 - TestRu	88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 4ae72d 4ae72d	B keras B keras	None None None None None None None None	None None None None None None None None	None None None None None None None None	0.812 0.777 0.81 0.802 0.821 0.815 0.812 0.651 0.82 0.818 0.508 0.812	0.405 0.452 0.403 0.415 0.387 0.399 0.408 0.609 0.389 0.391 0.693 0.404	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	l	opt_name× use_multiprocee opt_amsgrad× shuffle× opt_ opt_beta_1× steps_per_epoc validation_splity sample_weighty workers× bat validation_steps initial_epoch× max_queue_siz	monitor× earning_rate× m× patience> epochs× th_size× x opt_decay	< Adam —		0.02500 0.02500 0.02000 0.02000	90 0.990081718.0000 98000 11.000 9666990.00090 9000.00000 9000.00000 9000.00000 90000000 900000 900000 900000 900000 900000 900000 9000000 9000000 900000000	BL.00001 1.00000 00 -1.00000 00 -0.50000 - 00 -0.50000 -	2.2000 1.10000 2.20000 -1.10000 2.15000 - 2.10000 -1.05000 - 2.05000 - 2.05000 - 1.95000 -	195.00001.0 1.0000 150.0000 0.5000	0000 1.0000 11.000 1.00000 1.00000 7 0.50000 10.50000 - 0.50000 10.50000 -	000 1.00000 1.00000 -0.000 0.50000 -0.000	0.00000 1.09690 1.10 00 -1.10000 - 00 -1.05000 -1.05000 - 00 -1.05000 -1.05000 -		0.69116 0.65000 - 0.60000 -	0.65 0.6 0.55 0.5 0.4
	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	atlprd56 — TestRu atlprd56 — TestRu	88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 4ae72d 4ae72d 4ae72d	& keras	None None None None None None None None	None None None None None None None None	None None None None None None None None	0.812 0.777 0.81 0.802 0.821 0.815 0.812 0.651 0.82 0.818 0.508 0.812 0.802	0.405 0.452 0.403 0.415 0.387 0.399 0.408 0.609 0.389 0.389 0.391 0.693 0.404	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	l	opt_nameX use_multiproces opt_amsgradX shuffleX opt_ opt_beta_1X steps_per_epoc validation_splitX sample_weightX workersX bat validation_steps inittal_epochX max_queue_siz class_weightX	monitor× earning_rate× earning_rate× patience> epochs× th_size× × opt_decay min_delta×	< Adam —		0.025 0.0250000000000	90 0.90001718.0000 9000 1.10000 946000 0.0000 94600 1.0000 9000.6000 9000.6000 9000.6000 9000.6000 9	EL 00000 1.00000 00 -1.00000 - 00 -0.50000 - 00 -0.50000 - 00 -0.50000 - 00 -0.50000 - 00 -1.00000 -	2.2000 1.1000 2.2000 1.1000 1.0000 1.0000 2.0000 1.0000 2.0000 1.0000 1.0000 1.0000 1.9000 0.9000 N 1.5000 1.0000 -	195.00001.0 195.0000 150.0000 0.3000 www.th-eccol 0.0000 50.0000 V 10000	0000 1.0000 11.00 1.0000 1.0000 7 0.50000 10.50000 7 0.50000 10.50000 - 0.50000 9.50000 - 1.00000 9.50000 - 1.00000 9.00000 -	000 1.00000 1.00000 0.000 	0.00000 1.09490 1.10 00 1.05000 1.05490 1.10 00 1.05000 1.05000 1 00 1.05000 1.05000 1 00 0.95000 0.049888 00 0.95000 0.049800 1 00 0.95000 0.049000 1	-False - None - N	0.69116 0.65000 - 0.60000 - 0.50000 - 0.50000 - 0.40000 - 0.40000	0.6
	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	atiprd56 - TestRu atiprd56 - TestRu	88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 48a72d 4ae72d 4ae72d 4ae72d	B keras	None None None None None None None None	None None None None None None None None	None None None None None None None None	0.812 0.777 0.81 0.802 0.821 0.815 0.812 0.651 0.82 0.818 0.508 0.508 0.812 0.802 0.824	0.405 0.452 0.403 0.415 0.387 0.399 0.408 0.609 0.389 0.391 0.693 0.404 0.421 0.382		l	opt_nameX use_multiproces opt_amsgradX shuffleX opt_ opt_beta_1X steps_per_epoc validation_split sample_weightX workersX bat validation_stepp initial_epochX max_queue_siz class_weightX opt_epsilonX	monitor× earning_rate× n× patience> epochs× th_size× × opt_decay min_delta× opt_beta_2×	< Adam —		0.025 0.0250000000000	90 0.990081718.0000 98000 11.000 9666990.00090 9000.00000 9000.00000 9000.00000 90000000 900000 900000 900000 900000 900000 900000 9000000 9000000 900000000	EL 00000 1.00000 00 -1.00000 - 00 -0.50000 - 00 -0.50000 - 00 -0.50000 - 00 -0.50000 - 00 -1.00000 -	2.2000 1.1000 2.2000 1.1000 1 2.1500 1 2.5500 1 2.5500 1 2.5500 1 1.9500 0 1.9500 0 1.5500 1	195.00001.0 195.0000 150.0000 0.3000 www.th-eccol 0.0000 50.0000 V 10000	0000 1.0000 11.00 -1.00000 11.00000 - -0.50000 10.50000 - -0.50000 -9.50000 -	000 1.00000 1.00000 0.000 	0.00000 1.09490 1.10 00 1.105000 1.05490 1.10 00 1.05000 1.05000 0 00 1.05000 1.05000 0 00 0.95000 0.95000 0	-False - None - N	0.69116 0.65000 - 0.60000 -	0.6 0.55 0.5
	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	atlprd56 - TestRu atlprd56 - TestRu	88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 4ae72d 4ae72d 4ae72d 4ae72d 4ae72d	B keras	None None None None None None None None	None None None None None None None None	None None None None None None None None	0.812 0.777 0.81 0.802 0.821 0.815 0.812 0.651 0.812 0.818 0.82 0.818 0.802 0.824 0.822	0.405 0.452 0.403 0.415 0.399 0.408 0.609 0.389 0.391 0.693 0.404 0.421 0.382 0.385		l	opt_nameX use_multiproces opt_amsgrad× shuffle× opt_ opt_beta_1× steps_per_epoc validation_split> sample_weight\ workers× bat validation_steps initial_epoch× opt_epsilon× opt_epsilon×	monitor× earning_rate× n× patience> epochs× th_size× x opt_decay min_delta× opt_beta_2×	< Adam —		0.025 0.0250000000000	90 0.90001718.0000 9000 1.10000 946000 0.0000 94600 1.0000 9000.6000 9000.6000 9000.6000 9000.6000 9	EL 00000 1.00000 00 -1.00000 - 00 -0.50000 - 00 -0.50000 - 00 -0.50000 - 00 -0.50000 - 00 -1.00000 -	2.2000 1.1000 2.2000 1.1000 1.0000 1.0000 2.0000 1.0000 2.0000 1.0000 1.0000 1.0000 1.9000 0.9000 N 1.5000 1.0000 -	195.00001.0 195.0000 150.0000 0.3000 www.th-eccol 0.0000 50.0000 V 10000	0000 1.0000 11.00 1.0000 1.0000 7 0.50000 10.50000 7 0.50000 10.50000 - 0.50000 9.50000 - 1.00000 9.50000 - 1.00000 9.00000 -	000 1.00000 1.00000 0.000 -0.000 0.000 -0.0000 0.000 -0.0000 0.000 -1.00000 0.000	0.00000 1.09490 1.10 00 1.05000 1.05490 1.10 00 1.05000 1.05000 1 00 1.05000 1.05000 1 00 0.95000 0.049888 00 0.95000 0.049800 1 00 0.95000 0.049000 1	-False - None - N	0.69116 0.65000 - 0.60000 - 0.50000 - 0.50000 - 0.40000 - 0.40000	0.6
	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	atlprd56 — TestRu atlprd56 — TestRu	88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 4ae72d 4ae72d 4ae72d 4ae72d 4ae72d 4ae72d	B keras	None None None None None None None None	None None None None None None None None	None None None None None None None None	0.812 0.777 0.81 0.802 0.821 0.815 0.812 0.651 0.812 0.818 0.508 0.812 0.802 0.822 0.808	0.405 0.452 0.403 0.415 0.387 0.399 0.408 0.609 0.391 0.693 0.404 0.421 0.385 0.41			opt_nameX use_multiproces opt_amsgradX shuffleX opt_ opt_beta_1X steps_per_epoor validation_split validation_steps initia_epochX validation_steps initia_epochX class_weightX opt_epsilonX restore_best_wit	monitor× earning_rate× n× patience> epochs× th_size× x opt_decay min_delta× opt_beta_2×	< Adam —		0.025 0.0250000000000	90 0.90001718.0000 9000 1.10000 946000 0.0000 94600 1.0000 9000.6000 9000.6000 9000.6000 9000.6000 9	EL 00000 1.00000 00 -1.00000 - 00 -0.50000 - 00 -0.50000 - 00 -0.50000 - 00 -0.50000 - 00 -1.00000 -	2.2000 1.1000 2.2000 1.1000 1.0000 1.0000 2.0000 1.0000 2.0000 1.0000 1.0000 1.0000 1.9000 0.9000 N 1.5000 1.0000 -	195.00001.0 195.0000 150.0000 0.3000 www.th-eccol 0.0000 50.0000 V 10000	0000 1.0000 11.00 1.0000 1.0000 7 0.50000 10.50000 7 0.50000 10.50000 - 0.50000 9.50000 - 1.00000 9.50000 - 1.00000 9.00000 -	000 1.00000 1.00000 0.000 -0.000 0.000 -0.0000 0.000 -0.0000 0.000 -1.00000 0.000	0.00000 1.09490 1.10 00 1.05000 1.05490 1.10 00 1.05000 1.05000 1 00 1.05000 1.05000 1 00 0.95000 0.049888 00 0.95000 0.049800 1 00 0.95000 0.049000 1	-False - None - N	0.69116 0.65000 - 0.60000 - 0.50000 - 0.50000 - 0.40000 - 0.40000	0.6
	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	atiprd56 — TestRu atiprd56 — TestRu	88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 48a72d 4ae72d 4ae72d 4ae72d 4ae72d 4ae72d 4ae72d	B keras B keras	None None None None None None None None	None None None None None None None None	None None None None None None None None	0.812 0.777 0.81 0.802 0.821 0.815 0.651 0.822 0.818 0.82 0.818 0.812 0.802 0.824 0.822 0.824 0.802	0.405 0.452 0.403 0.415 0.387 0.399 0.408 0.609 0.391 0.693 0.404 0.421 0.382 0.385 0.404 0.321 0.382			opt_nameX use_multiproces opt_amsgrad× shuffleX opt_ opt_beta_1x steps_per_epoc validation_steps initial_epoch× max_queue_siz class_weight× opt_epsion× validation_rteq> restore_best_w baseline×	monitor× earning_rate× m× patience> epochs× h, size× x opt_decay min_delta× opt_beta_2× ights×	< Adam —		0.025 0.0250000000000	90 0.90001718.0000 9000 1.10000 946000 0.0000 94600 1.0000 9000.6000 9000.6000 9000.6000 9000.6000 9	EL 00000 1.00000 00 -1.00000 - 00 -0.50000 - 00 -0.50000 - 00 -0.50000 - 00 -0.50000 - 00 -1.00000 -	2.2000 1.1000 2.2000 1.1000 1.0000 1.0000 2.0000 1.0000 2.0000 1.0000 1.0000 1.0000 1.9000 0.9000 N 1.5000 1.0000 -	195.00001.0 195.0000 150.0000 0.3000 www.th-eccol 0.0000 50.0000 V 10000	0000 1.0000 11.00 1.0000 1.0000 7 0.50000 10.50000 7 0.50000 10.50000 - 0.50000 9.50000 - 1.00000 9.50000 - 1.00000 9.00000 -	000 1.00000 1.00000 0.000 -0.000 0.000 -0.0000 0.000 -0.0000 0.000 -1.00000 0.000	0.00000 1.09490 1.10 00 1.05000 1.05490 1.10 00 1.05000 1.05000 1 00 1.05000 1.05000 1 00 0.95000 0.049888 00 0.95000 0.049800 1 00 0.95000 0.049000 1	-False - None - N	0.69116 0.65000 - 0.60000 - 0.50000 - 0.50000 - 0.40000 - 0.40000	0.6
	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	atiprd56 — TestRu atiprd56 — TestRu	88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 88afbd 4ae72d 4ae72d 4ae72d 4ae72d 4ae72d 4ae72d 4ae72d 4ae72d	B keras	None None None None None None None None	None None None None None None None None	None None None None None None None None	0.812 0.777 0.81 0.802 0.821 0.815 0.812 0.651 0.812 0.818 0.508 0.812 0.802 0.822 0.808	0.405 0.452 0.403 0.415 0.387 0.399 0.408 0.609 0.391 0.693 0.404 0.421 0.385 0.41			opt_nameX use_multiproces opt_amsgradX shuffleX opt_ opt_beta_1X steps_per_epoor validation_split validation_steps initia_epochX validation_steps initia_epochX class_weightX opt_epsilonX restore_best_wit	monitor× earning_rate× m× patience> epochs× h, size× x opt_decay min_delta× opt_beta_2× ights×	< Adam —		0.025 0.0250000000000	90 0.90001718.0000 9000 1.10000 946000 0.0000 94600 1.0000 9000.6000 9000.6000 9000.6000 9000.6000 9	EL 00000 1.00000 00 -1.00000 - 00 -0.50000 - 00 -0.50000 - 00 -0.50000 - 00 -0.50000 - 00 -1.00000 -	2.2000 1.1000 2.2000 1.1000 1.0000 1.0000 2.0000 1.0000 2.0000 1.0000 1.0000 1.0000 1.9000 0.9000 N 1.5000 1.0000 -	195.00001.0 195.0000 150.0000 0.3000 www.th-eccol 0.0000 50.0000 V 10000	0000 1.0000 11.00 1.0000 1.0000 7 0.50000 10.50000 7 0.50000 10.50000 - 0.50000 9.50000 - 1.00000 9.50000 - 1.00000 9.00000 -	000 1.00000 1.00000 0.000 	0.00000 1.09490 1.10 00 1.05000 1.05490 1.10 00 1.05000 1.05000 1 00 1.05000 1.05000 1 00 0.95000 0.049888 00 0.95000 0.049800 1 00 0.95000 0.049000 1	-False - None - N	0.69116 0.65000 - 0.60000 - 0.50000 - 0.50000 - 0.40000 - 0.40000	0.6

IML Machine Learning

Summary (1)

- ✤ Goal of the project is to provide an ML/HPO service at ATLAS
 - To fulfil the demands that are expected from ML-topical physics, in particular in HL-LHC when larger dataset comes
 - An ATLAS directed work, but could be used by others as well
- The HPO workflow is developed and tested
 - "Ask" and "tell" are separated such that they are incorporated into ATLAS PanDA system
 - Docker container is used to preserve rapidly changing ML libraries
 - Tested with several use cases in ATLAS
 - Documented what need to change for a user with a mature training code

Summary (2)

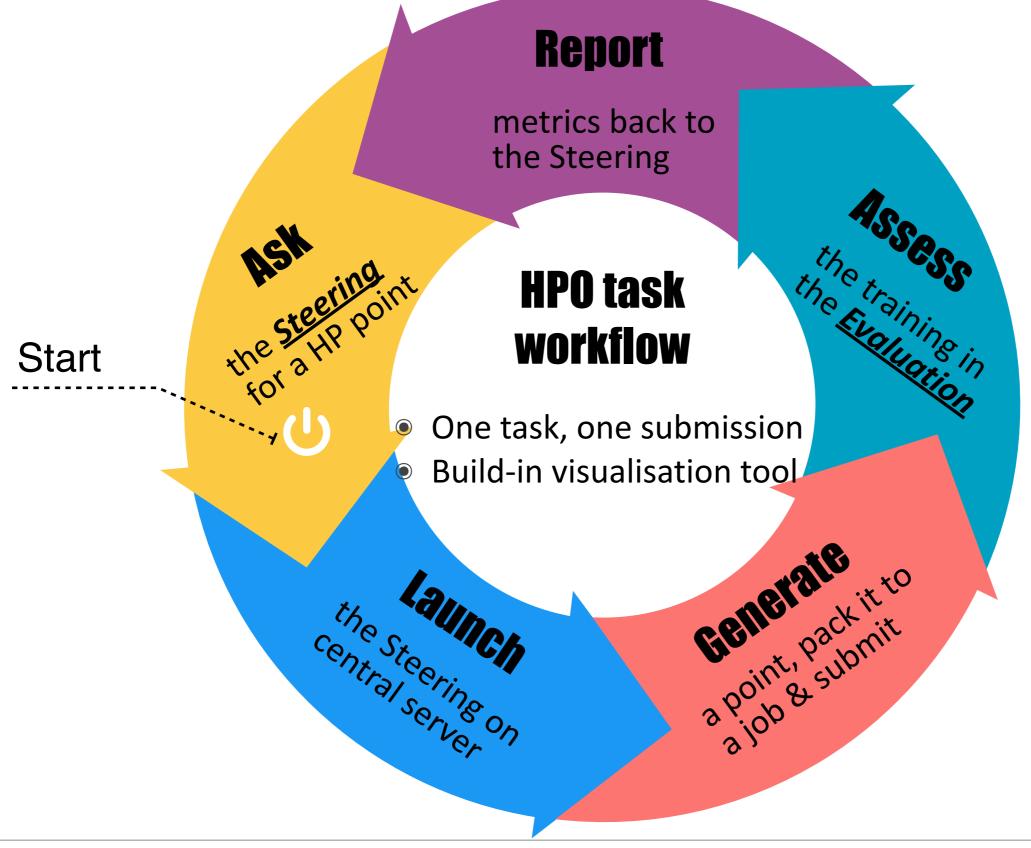
- Special scenarios
 - Distributed training
 - Some challenges were faced on Summit
 - Commercial Cloud (AWS) is currently being R&D
 - Segmented HPO
 - Implemented and tested with success
 - Convenient framework to train hundreds of models in one go; quite useful if ATLAS migrates to ML-based simulation
- Visualisation being supported centrally via the PanDA Monitoring system
 - Many new ideas can be implemented

Backup

Documentations

- Walk-through the Calo Image-based DNN example
 - SteeringContainer: <u>https://gitlab.cern.ch/zhangruihpc/</u> <u>SteeringContainer</u>
 - EvaluationContainer: <u>https://gitlab.cern.ch/zhangruihpc/</u> <u>EvaluationContainer</u>
- How to submit HPO task
 - <u>https://twiki.cern.ch/twiki/bin/view/PanDA/PandaHPO</u>
- iDDS Readme about the interfaces of ask-and-tell pattern
 - <u>https://idds.readthedocs.io/en/latest/usecases/</u>
 <u>hyperparemeter_optimization.html</u>

The HPO workflow



Steering container

- Run on central servers
- One container for all users
- Rich optimization algorithms
- Unified search space format

Evaluation container

- Run on Grid sites
- Encapsulate training job
- User customizable

ASSESE

Accessible to data on Grid

the training in Evolution the steering poin the steering point One task, one submission

checkpointing

- In case of short walltime on sites
- Periodically upload checkpoints to Grid igodol
- Download the checkpoint when retrying \bigcirc
- Resume training if checkpoint is found

More references

- Summit/HPC distributed training
- ✤ <u>New workflows for HPC</u>
- Summit/HPC challenges
- ✤ <u>New workflows</u>
- <u>4th Inter-experiment Machine Learning Workshop</u>
- Future analysis facility
- ✤ <u>AI for Big data</u>